

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 05002603 A

(43) Date of publication of application: 08.01.93

(51) Int. Cl

G06F 15/38

(21) Application number: 03147728

(71) Applicant: TOSHIBA CORP

(22) Date of filing: 20.06.91

(72) Inventor: TANABE NOBORU

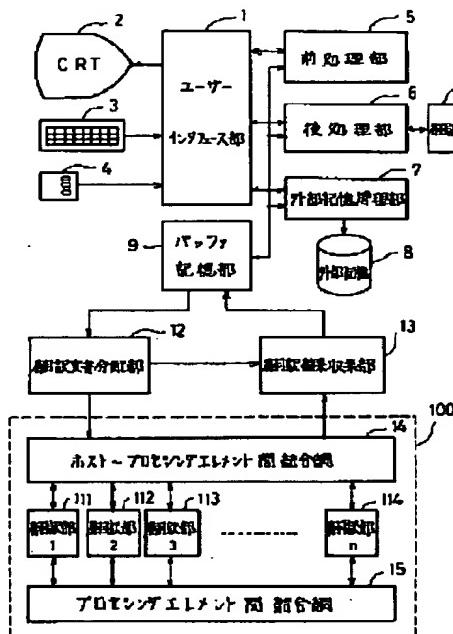
(54) PARALLEL MACHINE TRANSLATION DEVICE

(57) Abstract:

PURPOSE: To supply a parallel machine translation device in which the processing speed of a machine translation system is rapidly improved.

CONSTITUTION: Translation parts 111-114 arranged on plural processing elements which can be operated by an MIMD system, a data distribution part 12 which divides a document being the object of translation into plural parts and which distributes them to the respective processing elements, and a result collection part 13 which collects the parts of the document translated in the translation parts of the respective processing elements and which reedit them as the translation result of the original document and an interface part 1 controlling said respective parts and terminals 2-4 are provided. The respective partial documents are translated in parallel.

COPYRIGHT: (C)1993,JPO&Japio



HEI 5-2603

[ABSTRACT]

[PURPOSE]

To provide a parallel mechanical translation apparatus to drastically improve the processing speed of a mechanical translation system.

[CONSTITUTION]

A parallel mechanical translation apparatus for executing parallel translation of each partial document comprises the translation units 111 to 114 allocated on each element of a plurality of processing elements that may be operated in the MIMD system, a data distributing unit 12 for distributing the translation object document to each processing element by dividing it into a plurality of portions, a result collecting unit 13 for collecting the portions of the document translated in the translation unit for each processing element and editing again such portions as the translation result of the original document and an interface unit 1 for controlling the processes between each unit explained above and the terminals 2 to 4.

[0018]

Fig. 1 is a block diagram of the parallel mechanical translation apparatus as the first embodiment of the present invention. A user interface unit 1 displays a menu, documents and translation result on the CRT2

and accepts the operator's instruction inputted from a keyboard 3 and a mouse 4. The instructions accepted here are transferred to each portion such as pre-processor 5, a post-processor 6 and an external memory management unit 7 and are then executed.

[0019]

The characteristic items in adaptation of the present invention are that a plurality of translation units are allocated on the processing element of the MIMD type parallel computer (111 to 114) and these translation units respectively comprise a translated document distributing unit 12 for distributing the translation object documents and a translation result collecting unit 13 for connecting the result of translation on the processing element.

[0020]

An operator first designates a document to be translated, for example, a document written in the Japanese and then fetches this document into a buffer memory 9 via the external memory management unit 7. Next, the operator starts the pre-processor 5 to execute the adequate pre-processing for the translation unit to execute the translation.

[0021]

The document having completed the pre-processing is then transferred to a translated document distributing unit 12 based on the instruction from the

operator. The translated document distribution unit divides the translation object document into the sentences or paragraphs and then transmits these to the processing elements 111 to 114 via a coupling network 14 between the host and processing elements. In this case, respective sentences or paragraphs are given the number of identifying the positions in the original document. The information about the way of division is transferred to the translation result collecting unit 13 for the subsequent use to collect the result.

[0022]

Division of the translation object document is performed based on the relationship between the amount of document of translation object and the number of processing elements which may be used. For example, in the case where the number of processing elements 1000, number of sentences is 2000 and the number of paragraphs is 100, if the document is divided in unit of the paragraphs, 900 processing elements not used. However, the document is divided in unit of the sentence, the maximum parallel processing efficiency can be obtained because all processing elements can process two sentences. Moreover, when the document is divided in unit of sentence, a load can be adjusted in the finer unit than that in division in unit of the paragraph. Therefore, it is preferable also from the viewpoint of the unification of load. However, amount of

information to be processed in the distributing unit and collecting unit naturally increases.

[0023]

On the contrary, when the number of processing elements is 20 and the same document is divided with these processing elements, if the document is divided in unit of the paragraph, the maximum parallel processing efficiency can also be attained because all processing elements can process the five paragraphs. In addition, amount of information can be reduced more than that in the case where the document is managed in unit of sentence and distribution/collection overhead can also be reduced. Moreover, division in unit of paragraph little uses the dependence among the separated paragraphs even if high level translation using the preceding or succeeding sentences is executed and therefore communication between the processing elements can be eliminated.

[0024]

The processing element translates the assigned sentences to the designated language, for example, to the English from the Japanese. In the case where the distributed memory type MIMD parallel computer consisting of n processing elements is used, n translation units in maximum operate in parallel. Since the processing element of the distributed memory type MIMD type can be operated as one independent computer,

it is possible for the translation unit to use almost
indirect the translation algorithm which has been used
in the prior art.

[0025]

Since the translation is executed sentence by sentence in the existing translation algorithm, communication between the processing elements for translation can be eliminated. Therefore, the throughput of translation is improved almost to n times. It means, for example, that translation of 1000 sentences can be processed within about the transmission time of only one sentence using the parallel computer including 100 processors.

[0026]

In the existing translation system which has been used for practical processes, the program capacity is determined as about 30 Mbytes or so even when the user interface and dictionary are included. Therefore, when the parallel computer has the local memory of about 32 Mbytes, it is now possible to hold the translation unit including the dictionary on the processing element.

[0027]

Thereby, communications generated in the processing elements is only the input of the translation object document and output of the translation result and therefore the risk for resulting in the performance neck with input/output processes can be reduced. In

current, it is rather difficult to give local memories of 32 Mbytes to the parallel computer in the scale of 1000 units but it is sufficiently possible to realize such computer because 16 chips may be mounted when the 16 Mbytes DRAM can be utilized.

[0028]

Upon completion the translation, the processing element gives the identification number of the sentence before the translation to the result and then transmits the result with identification number to the translation result collecting unit 13 via the coupling network 14 between the host and processing element. Even when the processing elements start simultaneously the processes in the equal performance, the translation time is different for each sentence and therefore the translation result does not always reach the translation result collecting unit in the sequence of the identification number. Therefore, the translation result collecting unit buffers the result message to re-arrange the messages in the sequence of identification number and then transfers the result messages to the post-processor 6.

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平5-2603

(43)公開日 平成5年(1993)1月8日

(51)Int.Cl.⁵

識別記号 庁内整理番号

F I

技術表示箇所

G 06 F 15/38

T 9194-5L

審査請求 未請求 請求項の数5(全5頁)

(21)出願番号 特願平3-147728

(22)出願日 平成3年(1991)6月20日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 田邊 昇

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内

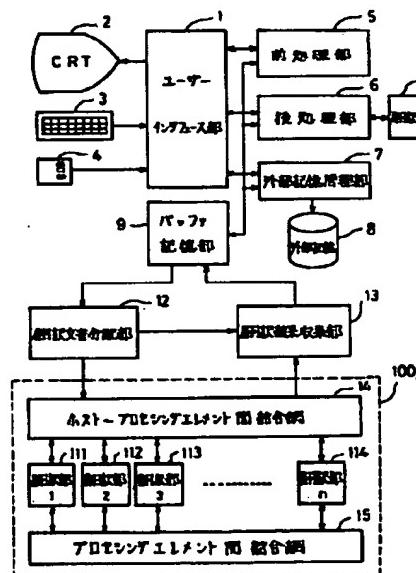
(74)代理人 弁理士 則近 憲佑

(54)【発明の名称】並列機械翻訳装置

(57)【要約】

【目的】機械翻訳システムの処理速度を飛躍的に向上させる並列機械翻訳装置の提供を目的とする。

【構成】MIMD方式で動作可能な複数のプロセシングエレメント各々上に配置される翻訳部111～114と、翻訳対象文書を複数の部分に分けて各プロセシングエレメントに分配するデータ分配部12と、各プロセシングエレメントの翻訳部で翻訳された文書の部分を収集してもとの文書の翻訳結果として再編成する結果收集部13と、以上の各部と端末2～4との間の制御を行なうインターフェース部1とを具備し、部分文書ごとに並列翻訳を行う並列機械翻訳装置である。



【特許請求の範囲】

【請求項1】 入力される文書を所定の文書に翻訳する複数のプロセシングエレメントと、これら複数のプロセシングエレメントに翻訳対象文書を分配するデータ分配手段と、このデータ分配手段により分配され各プロセシングエレメントにおいて翻訳された翻訳結果を収集する結果収集手段とを具備し、前記翻訳対象文書を並列に翻訳を行うことを特徴とする並列機械翻訳装置。

【請求項2】 請求項1に記載の並列機械翻訳装置において1個のプロセシングエレメントが担当する文書を、1個以上の文とする特徴とする並列機械翻訳装置。

【請求項3】 請求項1に記載の並列機械翻訳装置において1個のプロセシングエレメントが担当する文書を、1個以上の段落とする特徴とする並列機械翻訳装置。

【請求項4】 請求項1に記載の並列機械翻訳装置において1個以上の文を翻訳するために必要な部分を各プロセシングエレメントのメモリ上に保持することを特徴とする並列機械翻訳装置。

【請求項5】 請求項1に記載の並列機械翻訳装置において各プロセシングエレメントが担当する文の前後の部分をプロセシングエレメントが保持することを特徴とする並列機械翻訳装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は日本語や英語などの自然言語の間の相互の翻訳を支援する機械システムに関する。

【0002】

【従来の技術】 社会の国際化、情報化の進展、技術の進歩に伴うマニュアル、論文などの技術文書の増大は年々増加しており、翻訳のスループットを向上する技術の重要さが増してきている。

【0003】 従来、機械翻訳システムは汎用大型計算機やエンジニアリングワークステーションのような逐次型計算機上に実現してきた。しかし、その翻訳速度は1文あたり十秒から数十秒と必ずしも満足な速度であるとはいえない。

【0004】 このため、処理時間の関係から解析処理の全可能性を調べ尽くさず、適当なところで打ち切って後で帳尻を合わせるようなことで、翻訳の質を犠牲にするような弊害も引き起こしていた。さらに翻訳の質の向上のために従来以上に計算時間のかかる手法をとりいれることは困難であった。

【0005】 比較的安価に処理速度を向上させるために有力な方法に並列処理の導入が考えられるが、これまで実用化してきた機械翻訳システムは形態素解析や構文解析等の複数の段階に分かれており、たとえその一部を並列化しても、それだけでは全体の処理速度

は高々2倍程度にしかならない。それぞれの段階が複雑であるため、これらの全てを並列化することは極めて困難である。このようなアプローチは翻訳アルゴリズムの改良のたびに並列化を検討しなおさねばならないため、安易に改良を加えられなくなる原因となる。

【0006】 また機械翻訳アルゴリズムに内在する並列度は高々10程度しかないというのが定説であり、これまで並列処理によって機械翻訳を飛躍的に高速化することは困難であるといわれてきた。

10 【0007】 そこでこれまでに実用化されてきた手法とは根本的に異なるコネクショニストモデルのような方法で並列処理を自然言語処理に取り入れようという研究もある。ところがこの研究は機械翻訳システムとして実用化されるには至っておらず、これまで実績のある手法の完成度まで追いつく見通しはそれほど明るくない。また、この手法は翻訳の質の向上には寄与しても、速度の向上には必ずしもつながらない。

【0008】

【発明が解決しようとする課題】 このように、従来実用化されてきた逐次型計算機上の機械翻訳システムの処理速度は遅いという欠点があった。このため速度上の問題から翻訳の質の向上のために、より計算時間のかかる手法を導入することが困難であった。構文解析部の並列化のように部分的な並列化は全体の処理速度の飛躍的向上は望めないという欠点があった。従来実用化されてきた手法とは根本的に違うコネクショニストモデル等の手法では、従来実用化されてきたシステムの上に積み重ねられてきたノウハウが生かしにくいという欠点があった。本発明はこれまで並列化が困難とされてきた従来型の機械翻訳システムの上に積み重ねられてきたノウハウを殆どそのまま利用しながら、並列処理による飛躍的高速化を実現する手法を提供する。

【0009】

【課題を解決するための手段】 第一の発明では、MIMD (multiple instruction multiple data stream) 方式で、動作可能な複数のプロセシングエレメントと、これらのプロセシングエレメントと操作者のインターフェースをとる端末と、翻訳対象文書を各プロセシングエレメントに分配するデータ分配部と、プロセシングエレメント上に配置される分配された文書の翻訳部と、プロセシングエレメントから送信される部分翻訳結果を収集する結果収集部を具備させ、部分文書ごとに並列に翻訳を行わせる。第二の発明では、1個のプロセシングエレメントが担当する文書を、1個以上の文とする。第三の発明では、1個のプロセシングエレメントが担当する文書を、1個以上の段落とする。第四の発明では、1個以上の文を翻訳するために必要な部分を各プロセシングエレメントのメモリ上に保持する。第五の発明では、各プロセシングエレメントが担当する文の前後の部分をプロセシングエレメントが保持する。

【0010】

【作用】これまでに実用化されてきた機械翻訳システムは、ほとんど文相互の依存関係を用いずに基本的には文単位で独立に処理されてきた。そこで用いられてきた翻訳部（ユーザーインターフェース部などを除いた部分）をMIMD方式の並列計算機の各プロセシングエレメント上で動作させる。

【0011】MIMD方式の並列計算機ではSIMD (single instruction multiple datastream) 方式と異なり、他のプロセシングエレメントと全く同じ動作をしなければならないという制限がないため、プロセシングエレメント上の翻訳部は基本的には他のプロセシングエレメントとは完全に独立に動作でき、文ごとに動作の異なる従来型の翻訳部が並列に動作可能である。

【0012】端末上ではユーザーインターフェース部や分配部や翻訳結果収集部が動作しており、操作者は端末を介して翻訳対象文書や処理内容を指定し、翻訳結果収集部によって結果を検査したり、ディスクに格納したりする。

【0013】分配部はプロセシングエレメントの個数と、翻訳対象文書の量を考慮して、文や段落をプロセシングエレメントに分配する。プロセッサ数が文の数に比べて少ない場合は段落のように大きなまとまった量の文を分配する。こうするとデータ収集部の負担が軽減し、段落内の他の文を使った高度な翻訳を行うことも可能になる。逆にプロセッサ数が多い場合はプロセシングエレメントが担当する単位を分単位で細かくすることにより最大文書全体に含まれる文の個数の並列度が得られるのでスループットを飛躍的に向上できる。

【0014】プロセシングエレメントは自分に割り当てられた文を基本的には従来法に則って1文ずつ翻訳処理を行う。翻訳に際し他の文を用いていないのでプロセシングエレメント間での通信は無い。プロセシングエレメントでは独自に翻訳部が動いているのでこの処理は完全に並列に実行される。

【0015】1個以上の文を翻訳するために必要な部分を各プロセシングエレメントのメモリ上に保持する場合、ディスクアクセスも起こらないので並列化したときの性能が入出力で制限されるということもない。

【0016】プロセッサ数が多い場合、文単位で単純に並列処理すると他の文を用いた高度の翻訳をするためにプロセシングエレメント間の通信が生じ、通信オーバーヘッドが大きい計算機の場合は性能が低下してしまう。ところが翻訳する文の前後にある文を他のプロセシングエレメントと重複して予め保持しておけば翻訳中にプロセシングエレメント間通信を行わなくても、他の文を用いた高度な翻訳が可能になる。

【0017】

【実施例】以下図面を参照して本発明の実施例を説明する。

【0018】図1は本発明の第一の実施例の並列機械翻訳装置のブロック図である。ユーザーインターフェース部1はCRT2上にメニューや文書や翻訳結果などを表示し、キーボード3やマウス4から入力される操作者からの指示を受け付ける。ここで受け付けられた指示は前処理部5や後処理部6や外部記憶管理部7などの各部に伝えられ、実行される。

【0019】本発明の適用において特徴的であるのは、複数の翻訳部がMIMD型並列計算機のプロセシングエレメント上に配置され(111～114)、これらに翻訳対象文書を分配する翻訳文書分配部12と、プロセシングエレメント上で翻訳された結果を収集する翻訳結果収集部13が存在することである。

【0020】まず操作者は翻訳したい文書、例えば日本語で書かれたある文書を指定し、外部記憶管理部7を介してバッファ記憶部9に取り込む。次に翻訳部が取り扱えるように前処理部5を起動して適度に前処理を加える。

【0021】前処理の終わった文書は操作者の指示に基づき、翻訳文書分配部12に送られる。翻訳文書分配部では、翻訳対象文書を文または段落に分割してホスト～プロセシングエレメント間結合網14を介してプロセシングエレメント111～114に送信する。このときそれぞれの文または段落には元の文書中での位置を識別できる番号を付与しておく。どのように分割されたかという情報は翻訳結果収集部13に伝えられ、結果収集の際に用いられる。

【0022】翻訳対象文書の分割の仕方は、翻訳対象の文書の量と使用可能なプロセシングエレメントの個数の関係に基づいて行われる。たとえばプロセシングエレメント数が1000個で文の数が2000個、段落の数が100個だった場合は、段落の単位で分割してしまうと900個のプロセシングエレメントが遊んでしまう。しかし、文の単位で分割すれば全てのプロセシングエレメントが2個の文を担当できるので最大の並列度が得られる。また文単位で分割すれば段落単位で分割するよりも細かい単位で負荷を調整できるので、負荷の均一化の観点からも好ましい。しかし分配部や収集部が管理すべき情報量が増加する。

【0023】逆にプロセシングエレメント数が20個で同じ文書を分割するならば段落の単位で分割しても全てのプロセシングエレメントが5個の段落を担当できるので最大の並列度が得られる。さらに文単位で管理するよりは情報量が少なくてすみ、分配・収集オーバーヘッドも少なくできる。さらに段落単位で分割すれば前後の文を用いた高度な翻訳を行う場合でも、離れた段落間にまたがる依存関係を使うことは殆どないので、ほぼプロセシングエレメント間の通信を無くすことができる。

【0024】プロセシングエレメントでは割り当てられた文を指定された言語たとえば日本語から英語に翻訳す

る。n個のプロセシングエレメントからなる分散メモリ型MIMD型並列計算機を用いた場合、最大n個の翻訳部が並列に動作する。分散メモリ型MIMD型のプロセシングエレメントは1個の独立した計算機として動作できるので、翻訳部は従来から実用化されてきた翻訳アルゴリズムをほとんどそのまま使うことも可能である。

【0025】従来の翻訳アルゴリズムであれば文ごとに独立に翻訳されるので、翻訳のためにプロセシングエレメント間通信をしなくてよい。このため翻訳のスループットはほぼn倍に高速化する。たとえば1000台のプロセッサを持つ並列計算機を用いて1000個の文の翻訳を行うのに1個の文の翻訳時間程度で処理できることを意味する。

【0026】従来実用化されてきた翻訳システムではユーザーインターフェース部や辞書を含めてもプログラムの容量は30メガバイト前後と言われているので、32メガバイト程度のローカルメモリを持つ並列計算機であれば辞書を含めた翻訳部をプロセシングエレメント上に保持することが可能である。

【0027】このようにすればプロセシングエレメントで生じる通信は翻訳対象文書の入力と翻訳結果の出力のみとなり、入出力が性能のネックになる危険性は少ない。現時点では1000台規模の並列計算機で32メガバイトものローカルメモリを持たせるのはやや困難だが、16メガDRAMが利用できる頃には16チップで実装できるので十分実現可能である。

【0028】プロセシングエレメントは翻訳が終了すると翻訳前の文が持っていた識別番号を結果に付与し、ホスト～プロセシングエレメント間結合網14を介して翻訳結果収集部13に送信する。プロセシングエレメントが同一の性能で同時に処理を開始したとしても、文ごとに翻訳の時間は異なるので翻訳結果は必ずしも識別番号順には翻訳結果収集部には届かない。そこで翻訳結果収集部は結果メッセージをバッファリングし、番号順に並べ変えて後処理部6へ渡す。

【0029】操作者はユーザーインターフェース部1を介してその結果を見て、後処理部を用いて例えば英語の誤訳を直接英語でなおしたり、正しい結果が得やすいと思われるよう翻訳対象文を日本語で修正して再度上記の方法で翻訳を繰り返す。なお、後処理で変更箇所が少ない場合は並列計算機上の翻訳部ではなく端末側の翻訳10で対話的に行っても良い。また、あらたに修正が指示された部分だけ並列計算機側に送り、その結果を前回の結果を入れ換えてよい。

【0030】プロセシングエレメント上の翻訳部で従来の翻訳アルゴリズムを用いるのであればプロセシングエレメント間の通信はないが、他の文を用いた高度な翻訳をする場合、例えばプロセシングエレメント1個に1文

しか割り当てられていなかったりするとプロセシングエレメント間の通信が生じる。通信オーバーヘッドが大きい計算機の場合は性能が低下してしまう。ところが翻訳する文の前後にある文を他のプロセシングエレメントと重複して予め保持しておけば翻訳中にプロセシングエレメント間通信を行わなくとも、他の文を用いた高度な翻訳が可能になる。

【0031】以上の実施例では主に分散メモリ構成のMIMD型並列計算機を用いた場合について説明したが、プロセシングエレメント上で翻訳部が動作でき、1つ以上の端末からの指示でデータの分配・収集ができるシステムであれば、ネットワーク環境で相互に結合された複数のワークステーションやプロセッサプール（計算サーバ）からなるシステムや、共有メモリ構成のMIMD型並列計算機にも本発明は適用できる。また、本実施例ではユーザーインターフェース部や翻訳対象文書分配部や翻訳結果収集部がそれぞれ1つの場合を示したが複数あってよい。

【0032】また、プロセシングエレメント間通信は無いに等しいので、プロセシングエレメント間結合網15の結合形態は問わず、メッシュでも、トリーでも、イーサネットでも、あらゆる結合形態のシステムに適用できる。プロセシングエレメント間通信を皆無にできる場合は、プロセシングエレメント間結合網はなくてもよい。さらに、本発明では逐次型の計算機上で動作するいかなるアルゴリズムでも、記憶容量を越えない限り適用でき、並列計算機上に容易に移植可能である。

【0033】

【発明の効果】以上のように本発明によればこれまで並列化が困難とされてきた従来型の機械翻訳システムの上に積み重ねられてきたノウハウを殆どそのまま利用しながら、並列処理による機械翻訳の飛躍的高速化を実現できる。また文間にまたがる情報を利用した高度な翻訳を行う場合でも対応できる。さらに、処理の高速性を利用して従来より時間がかかるような翻訳アルゴリズムでも、並列計算機に容易に取り入れることができる。

【図面の簡単な説明】

【図1】 本発明の一実施例に係る並列機械翻訳装置の構成を示すブロック図。

【符号の説明】

- | | |
|----|--|
| 40 | 1…ユーザーインターフェース部, 2…CRT, 3…キーボード, 4…マウス, 5…前処理部, 6…後処理部, 7…外部記憶管理部, 8…外部記憶, 9…バッファ記憶部, 10…翻訳部, 111, 112, 113, 114…プロセシングエレメント（翻訳部）, 12…翻訳対象文書分配部, 13…翻訳結果収集部, 14…ホスト～プロセシングエレメント間結合網, 15…プロセシングエレメント間結合網, 100…並列計算機 |
|----|--|

【図1】

